COMPARISON OF THE STANFORD-BINET INTELLIGENCE SCALE WITH THE GOODENOUGH-HARRIS DRAW A PERSON AND THE DRAW A PERSON: A QUANTITATIVE SCORING SYSTEM

By

SHEILA MARIE KRAEMER

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1991

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
May, 1993

COMPARISON OF THE STANFORD-BINET
INTELLIGENCE SCALE WITH THE
GOODENOUGH-HARRIS DRAW A
PERSON AND THE DRAW A
PERSON: A QUANTI-
TATIVE SCORING
SYSTEM

Thesis Approved:

Keith E. Jones
Thesis Adviser

Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Ruth Tomes, for her encouragement and advice throughout my graduate program and especially during the completion of this project. The support and suggestions of Dr. Arlene Fulton and Dr. Patricia Self were also invaluable.

Without the encouragement of my mom and dad, Emma and Larry Heard, I would never have been able to complete this project. They have been a constant source of support and believed in me even when I didn't. For that I am eternally grateful. I also owe a special thanks to my uncle, Bill Oliver, and my aunt, Inez Chapman, for their continual words of praise and encouragement throughout my studies.

To my best friend, Stacy Thompson, who shared with me the ups and downs of graduate school, always there with an uplifting smile and contagiously cheerful attitude, goes my heart-felt thanks. I would also like to thank Jane Jacob for her calming words of encouragement and help in editing my thesis. To my husband Peter, for the long hours, constant support, and for never doubting I could do it, I am forever indebted. And most importantly, I thank God. For only by His grace was any of this possible.

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE PROBLEM

Introduction

Two newly revised measures of intelligence have recently been introduced into the field of assessment. In 1986, the new Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE) (Thorndike, Hagen, & Sattler, 1986) was released. In 1988, the Draw a Person: A Quantitative Scoring System (DAP) (Naglieri, 1988) was released. Both of these new tests claim to assess intelligence but are based on different theories of intelligence.

The SB:FE is a standardized IQ test based on a theory of general intelligence. The general intelligence is comprised of four factors or cognitive abilities which are placed in a three level hierarchical model. The SB:FE represents a significant change in theory from the early revisions. The early theories defined intelligence as a mental quotient which was determined by dividing a persons mental age (determined by testing) by their chronological age (Sattler, 1990). Earlier versions as well as the SB:FE use standard deviations in the assessment of intelligence.

The DAP is an intelligence test based on the drawing theories of intelligence passed down from the authors of the first widely accepted human figure drawing test developed by Goodenough (Naglieri, 1988). The DAP was developed as a response to criticisms of the Goodenough-Harris Draw a Person test (G-H) (Harris, 1963). It is the purpose of this test to provide a more valid and reliable measure of intelligence than its counterpart, the G-H.

The problem arises, though, as to whether the DAP can fulfill that purpose if it still embraces the old theories of drawing development rather than those of the traditional intelligence tests. A question exists as to whether the human figure drawings used as measures of intelligence assess the same constructs as standard intelligence tests.

Purpose

What is the relationship between children's human figure drawings and their intelligence? This study assesses that relationship with the following measures: the Goodenough-Harris Draw a Person (G-H) (Harris, 1963), Draw a Person: A Quantitative Scoring System (DAP) (Naglieri, 1988), and the Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE) (Thorndike, Hagen, & Sattler, 1986). Specifically, woman, man, and composite scores from both drawing tests are compared with area and composite scores of the SB:FE. Objectives of this research include:

1. To investigate the relationship between G-H scores and scores from the SB:FE, specifically:

(a) To assess relationships between the G-H man score and the set of SB:FE scores (Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, Short-Term Memory, and Composite).

(b) To assess relationships between the G-H woman score and the set of SB:FE scores.

(c) To assess relationships between the G-H composite score and the set of SB:FE scores.

2. To investigate the relationship between DAP scores and scores from the SB:FE, specifically:

(a) To assess relationships between the DAP man score and the set of SB:FE scores.

(b) To assess relationships between the DAP woman score and the set of SB:FE scores.

(c) To assess relationships between the DAP self score and the set of SB:FE scores.

(d) To assess relationships between the DAP composite score and the set of SB:FE scores.

3. To determine if one drawing test correlates significantly higher with the SB:FE Composite Score.

Importance of the Study

The first justification for this research is that results will add to the literature important information about theories of intelligence. Theoretically, it will render a better understanding of the concepts tapped by the two new intelligence tests and provide information concerning their construct and concurrent validity.

Secondly, it will render applied benefits as well. It is important for current psychologists and psychometrists to know the validity and reliability of the tests they use. This research will provide general information as to the usefulness of two of the newest assessment measures.

This thesis is presented in five chapters, the first of which has been the introduction of the problem. The second chapter consists of a review of the relevant research literature already conducted using the SB:FE, G-H, and DAP. The third chapter details the methods used in conducting the research. The fourth chapter presents the results and the fifth chapter summerizes the research with a discussion of the results, after which conclusions are drawn.

Definition of Terms

The following definitions will be used in the research:

1. DAP scores: The DAP man score, woman score, self score, and composite score are standard scores based on a mean of 100 and having a standard deviation of 15.

2. G-H scores: The G-H man score, woman score, and composite score are standard scores based on a mean of 100 and having a standard deviation of 15.

3. SB:FE area scores: The SB:FE Abstract/Visual score, Verbal Reasoning score, Quantitative score, and Short-Term Memory score are standard scores based on a mean of 100 and having a standard deviation of 16.

4. SB:FE Composite score: The SB:FE Composite score is a standard score based on a mean of 100 and having a standard deviation of 16.

CHAPTER II

LITERATURE REVIEW

Organization of Literature

The first section of the literature review will discuss the concept of intelligence and the legitimacy of using the newly revised SB:FE as a measure of this concept. Articles assessing the validity of this measure will be reviewed.

The second section of the literature review will look at the theory behind using human figure drawings as a measure of intelligence. Reviews of the validity and reliability of the G-H as a measure of intelligence will be presented first. Secondly, the current research using the new DAP scoring criteria will be reviewed.

The last section of this review will summarize and draw conclusions based on the literature presented in this chapter. At the end of the literature review hypotheses will be stated.

SB:FE and Measuring Intelligence

As mentioned earlier, the SB:FE is based on a theory of intelligence which incorporates general intelligence, better known as *g*, into a three level hierarchical model (Thorndike, Hagen, & Sattler, 1986). Charles Spearman was the first to introduce the idea of a general intelligence which he called the General Ability factor. He believed that by measuring *g* one could tap into all cognitive processes and assess a person's intelligence using only this one factor (Weinberg, 1989; Thorndike, 1990; Thorndike, Hagen, & Sattler, 1986). Although his theory was not embraced by all, modifications of *g* were incorporated into other researcher's theories, including those of Philip Vernon, Thomson, and Simpson (Thorndike, 1990). Another well known researcher, E.L. Thorndike, disagreed with Spearman and postulated instead that intelligence could not be interpreted or defined by one construct but was made up of many. He felt there were at least three different kinds of intelligence including abstract intelligence, social intelligence, and mechanical intelligence (Thorndike, 1990). L. L. Thurstone developed a multidimensional model to assess intelligence. He believed intelligence was made up of several different primary mental abilities and while he did not postulate that *g* existed, the correlation of the abilities he proposed did in effect produce a *g*. Also a multidimensionalist, J. P. Guilford developed a three-dimension model of intelligence (Thorndike, Hagen, & Sattler, 1986). John Horn and Raymond Cattell developed a model of "fluid and crystallized ability" which incorporated "nature and nurture" theories of development (Thorndike, Hagen, & Sattler, 1986). Fluid ability represents the biological givens while crystallized ability represents environmental experiences of one's life.

It is plain to see that many different theories of intelligence have been postulated throughout history and to this date no theory has been adopted as "truth". Researchers have found support for their individual theories and thus no one theory can be discarded for lack of evidence. The SB:FE, through validation studies reported in the technical manual (Thorndike, Hagen, & Sattler, 1986), has thus adopted a theory which incorporates the afore mentioned ideas. It incorporates *g* with Horn and Raymond's model of fluid and crystallized ability. It is postulated that *g* will stem from a set of cognitive tasks that comprise a host of contexts (Thorndike, Hagen, & Sattler, 1986). In the end a three level hierarchical model was established (refer to Figure 1, p. 30) with g at the top, crystallized abilities, fluid-analytic, and short-term memory on the second level, and verbal reasoning, quantitative reasoning, and abstract/visual reasoning on the third level (Thorndike, Hagen, & Sattler, 1986).

The SB:FE was chosen as the comparative intelligence test in this study due to its newer norms, general acceptance in the testing arena, and validity and reliability as reported in various articles. The SB:FE has been compared with the Wechsler Intelligence Scale for Children-Revised (WISC-R), the Kaufman Assessment Battery for Children (K-ABC), the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R), the Stanford-Binet L-M, the Woodcock-Johnson Tests of Achievement, and the Peabody Picture Vocabulary Test-Revised (PPVT-R).

Comparison of the SB:FE and WISC-R

Research comparing the SB:FE with the WISC-R yielded positive results overall with the exception of a study done by Carvajal and Weaver (1989). In their study comparisons were made between the SB:FE and WISC-R on a population of gifted children. The study selected 39 children enrolled in a gifted program as determined by scores on the WISC-R. Results indicated that the only significant correlation (p < .01) appeared between the SB:FE Verbal Reasoning area and the WISC-R Verbal IQ. This study contains at least two serious flaws. First, the SB:FE was given one to two years after the WISC-R had been administered. Secondly, several different psychologists gave the WISC-R while the SB:FE was administered by the researchers. Thus this study has difficulty establishing internal validity.

Phelps (1989) also did a study comparing the WISC-R with the SB:FE using a sample of gifted children. The WISC-R and the SB:FE were administered in counterbalanced order to a previously identified gifted population of students (n=48). Results showed that the SB:FE Composite significantly correlated (p<.01) with the WISC-R Full Scale IQ (r=.393). Area scores differed in their significance to WISC-R IQ Scales. The SB:FE produced lower scores on average, although not significantly, than the WISC-R.

Phelps, Bell, and Scott (1988) conducted a study comparing the WISC-R and the SB:FE on a population of learning disabled students. The SB:FE and the WISC-R were administered in counterbalanced order to a group of 35 previously classified learning disabled students. All correlations between scores on the SB:FE and WISC-R IQ scales were significant (ranging from $p<.05$ to .0001). The correlation between SB:FE Composite and WISC-R Full Scale IQ was the highest ($r=.92$).

The SB:FE and the WISC-R have also been compared on a population of learning disabled and developmentally handicapped children (Hollinger & Baldwin, 1990). The SB:FE and WISC-R were administered to a sample of 19 students identified as learning disabled or developmentally handicapped. Significant correlations were found between the SB:FE scores and WISC-R IQ's (ranging from $p<.05$ to .001). The only nonsignificant correlation came from WISC-R Performance Scale and SB:FE Verbal Reasoning.

The SB:FE and the WISC-R were administered to children labeled as functioning at the retarded level (n=30) in a study conducted by Lukens (1990). A Pearson r correlation of .83 ($p<.001$) was reported between the SB:FE Composite score and the WISC-R Full Scale IQ. The SB:FE Composite score correlated significantly at the .001 level with all three of the WISC-R IQ Scales. SB:FE Area scores were not reported.

Greene, Sapp, and Chissom (1990) conducted a study comparing the SB:FE and the WISC-R with a population of black students receiving special education services (n=51). Again it was reported that all correlations between scores on the SB:FE and the WISC-R IQ Scales were significant ($p<.001$).

Rothlisberg (1987) conducted a study to determine the concurrent validity of the SB:FE and the WISC-R. She chose a sample of 32 nonexceptional children. The tests were administered in counterbalanced order. Results of the study reported significant correlations between the two intelligence measures. Significance ranged from .05 to .001 across SB:FE Composite and Area scores with WISC-R IQ Scales. Two correlations did not reach significance: WISC-R Performance Scale with SB:FE Verbal Reasoning, and WISC-R Verbal Scale with SB:FE Quantitative Reasoning.

After reviewing research comparing the WISC-R with the SB:FE, convergent as well as concurrent validity seem to be established between these two measures of intelligence. Also it appears that the SB:FE is an adequate tool for determining the placement of children at different ends of the spectrum, including gifted, learning disabled, and mentally retarded. It also appears to be as good a measure as the WISC-R for assessing the intelligence of both black and white children.

Comparison of the SB:FE with the K-ABC

The SB:FE was compared with the K-ABC by Hendershott, Searight, Hatfield, and Rogers (1990). The two tests were administered in a counterbalanced order with a sample of 36 preschool children. Significant correlations were found between scores from the SB:FE and K-ABC scores ($p<.05$).

Only three correlations were not significant: K-ABC Simultaneous Processing with SB:FE Verbal Reasoning, K-ABC Simultaneous Processing with SB:FE Abstract/Visual Reasoning, and K-ABC Sequential Processing with SB:FE Quantitative Reasoning.

Another study comparing the SB:FE and K-ABC consisted of a sample of black learning-disabled students (Knight, Baker, & Minder 1990). Tests were administered in counterbalanced order to the students who had previously been identified as learning disabled. The Composite SB:FE and Composite K-ABC correlated significantly ($\underline{p}<.001$). Significant correlations were also found between the majority of other scores from the SB:FE and K-ABC ($\underline{p}<.05$ to .001). Only two correlations did not reach significance: the K-ABC Achievement with SB:FE Abstract/Visual, and the K-ABC Simultaneous Processing with SB:FE Quantitative Reasoning.

The K-ABC and the SB:FE were also compared using a sample of gifted students (Hayden, Furlong, and Linnemeyer, 1988). Tests were administered in counterbalanced order to a sample of 29 students enrolled in a gifted program. Eleven of the fifteen correlations calculated between scores from the SB:FE and the K-ABC were significant ($\underline{p}<.01$). The K-ABC Composite and the SB:FE Composite had the highest correlation ($\underline{r}=.70$).

Lamp and Krohn (1990) compared the SB:FE and the K-ABC with a sample of black and white children from low income families. Tests were administered in counterbalanced order to 39 white and 32 black children. Correlations were all significant across both populations and both tests ($p<.001$) with one exception. The correlation of SB:FE Quantitative Reasoning Area for black students did not reach significance ($r=.32$).

The SB:FE appears to be as good a measure as the K-ABC for measuring intelligence across populations. Again the SB:FE was able to differentiate between different populations including normal, learning disabled, and gifted. Evidence has also been provided to suggest that the SB:FE is an adequate measure of black children's intelligence.

Comparison of the SB:FE and the Stanford-Binet L-M

A study comparing the SB:FE, Stanford-Binet L-M: Third Edition, and the K-ABC was conducted by Krohn and Lamp (1989). The sample consisted of 89 head start students. The three tests were administered in counterbalanced order within a 2-week period during the last few weeks of the school year. Correlations were computed between the Stanford-Binet L-M IQ, SB:FE Area and Composite scores, and K-ABC Composite and Area scores. All correlations were significant ($p<.001$).

Lukens (1988) compared the SB:FE and the Stanford-Binet L-M using a sample of mentally retarded persons. The two tests were administered by the same examiner to 31 adolescents classified as mentally retarded. The Stanford-Binet L-M IQ was correlated with the SB:FE Composite and

Area scores. All correlations were reported as significant ($p<.001$, r ranged from .67 to .86).

A study conducted by Hartwig, Sapp, and Clayton (1987) compared the SB:FE with the Stanford-Binet L-M using a sample of 30 nonexceptional elementary students. One researcher administered both tests in a counterbalanced order. The correlation between Composite SB:FE and Stanford-Binet L-M IQ was significant ($p<.001$). Correlations ranged from moderate to high between the Stanford-Binet L-M IQ and the Area scores from the SB:FE (r ranged from .40 to .81).

Another comparative study between the SB:FE and Stanford-Binet L-M was conducted by Clark, Wortman, Warnock, and Swerdilik (1987). The tests were given in counterbalanced order to a sample consisting of 47 nonexceptional three to six year olds. A significant correlation was found between only two of the five reported. The Composite SB:FE with Stanford-Binet L-M IQ, and the Stanford-Binet L-M IQ with Short-Term Memory Area score reached significance ($p<.05$).

McCall, Yates, Hendricks, Turner and McNabb (1989) also conducted a study comparing the Stanford-Binet L-M with the new SB:FE. The subjects consisted of 32 children identified as gifted. The tests were administered in counterbalanced order, but the school psychologists administered the L-M version while authors of the study administered the SB:FE, calling into question the validity of the study. Low correlations were reported and none reached significance (*r* ranged from .13 to .28).

All but one study reviewed found positive, significant correlations between the SB:FE and the Stanford-Binet L-M. The SB:FE appears to be as good a measure of intelligence across populations (normal, mentally retarded, and disadvantaged) as the old Stanford-Binet L-M. The one study that reported no significance also reported a flawed methodology raising questions as to the validity of its results.

Comparison of the SB:FE With Other Measures

Carvajal, Parks, Bays, Logan, Lujano, Page, and Weaver (1991) looked at the relationships between scores on the WPPSI-R and the SB:FE. The sample consisted of 51 nonexceptional children from three to seven years of age. Trained examiners administered both tests. The three WPPSI-R scores were compared with the SB:FE Composite and Area scores. Twelve of the fifteen correlations calculated were significant ($p<.01$). Nonsignificant correlations were as follows: WPPSI-R Performance IQ with SB:FE Verbal Reasoning Area, WPPSI-R Performance IQ with SB:FE Quantitative Reasoning Area, WPPSI-R Verbal IQ with SB:FE Abstract/Visual Reasoning Area.

The SB:FE has also been compared to the WPPSI by Carvajal, Hardy, Smith, and Weaver (1988). A sample of 20 children were given the SB:FE and the WPPSI in counterbalanced order. Six correlations were calculated between the SB:FE and the WPPSI which resulted in only one significant correlation ($p<.01$). The Composite SB:FE and WPPSI Full Scale IQ were significantly correlated ($r=.587$).

Gifted children were tested by both the SB:FE and the PPVT-R in a study done by Carvajal (1988). Carvajal administered the PPVT-R first and then the SB:FE to a sample of 51 students identified as gifted. The PPVT-R score was compared with both the Composite and Area SB:FE scores. Although low, all correlations reached significance ($p<.05$ to .01).

In a study by Rothlisberg (1990) 31 grade school children were administered the SB:FE, the Wide Range Achievement Test-Revised (WRAT-R), and the Woodcock-Johnson Psychoeducational Battery: Tests of Achievement (WJTA). Correlations were significant between each achievement score and SB:FE scores with the following exceptions: The SB:FE Quantitative Reasoning Area score did not reach significance with any scores from the two achievement tests, the WRAT-R Arithmetic score did not correlate with the SB:FE Verbal Reasoning or Short-Term Memory Areas.

Although the SB:FE does not have the high correlations with the WPPSI and WPPSI-R as reported with other intelligence measures, it still appears to be measuring the same constructs. The relationship between the PPVT-R and SB:FE also validated the use of the SB:FE. An interesting finding does come forward when comparing the SB:FE with achievement tests. Although correlations were significant in the other areas, the Quantitative Reasoning area does not seem to measure the same constructs as do WRAT-R and WJTA.

SB:FE as an Adequate Measure of Intelligence

In reviewing the validity of the SB:FE it appears that both construct and concurrent validity have been established. The SB:FE seems to measure the same constructs as other widely accepted measures of intelligence. With its newer norms it may even be a better estimate than the older tests. It also did a good job of identifying different populations including the gifted (Carvajal, 1988; Hayden, Furlong, & Linnemeyer, 1988; Phelps, 1989), learning disabled (Knight, Baker, & Minder, 1990; Phelps, Bell, & Scott, 1988; Hollinger & Baldwin, 1990), mentally retarded (Lukens, 1988; Lukens, 1990), normal (Rothlisberg, 1990; Carvajal, Hardy, Smith, & Weaver, 1988; Carvajal, Parks, Bays, Logan, Lujano, Page, & Weaver, 1991; Clark, Wortman, Warnock, & Swerdlik, 1987; Hartwig, Sapp, & Clayton, 1987; Hendershott, Searight, Hatfield, & Rogers, 1990; Rothlisberg, 1987), and Head Start groups (Krohn & Lamp, 1989). Results also indicate that the SB:FE does not discriminate between black and white children (Lamp & Krohn, 1990). Laurent, Swerdlik and Ryburn (1992) reported similar findings. They concluded that the SB:FE was as good a measure of general intelligence as any other instrument testing intellectual ability.

Reliability of the SB:FE is quite good (Sattler, 1990), with internal consistency reliabilities for the Composite Score ranging from .95 to .99. Median reliabilities for subtests range from .73 to .94. Test-retest reliabilities have also been assessed. The stability coefficients for 5-year-olds and 8-year-olds are .91 and .90, respectively. Although some disagreement exits as to the true SB:FE factors (Laurent, Swerdlik, & Ryburn, 1992; Gridley & McIntosh, 1991; Ownby & Carmin, 1988; Thorndike, 1990; Kline 1989), conclusive evidence has not been produced which settles this aurgument or gives reason to reject the four factors of SB:FE.

Drawings as Measures of Intelligence

When one thinks of assessing intelligence by use of drawings the first name that seems to come to mind is Florence Goodenough. Her 1926 scale for assessing intelligence with human figure drawings was the first widely accepted and used of these types of scales. She based her scale on research which described the content of children's drawings as dependent on the child's intellectual functioning (Goodenough, 1926). One of the most extensive of these studies resulted in the identification of stages in the drawing of human figures by Rouma (Harris, 1963). There are basically two stages:

I. The preliminary stage

1. The child adapts his hand to the instrument.
2. The child gives a definite name to the

incoherent lines which he traces.

3. The child announces in advance what he intends to represent.
4. The child sees a resemblance between the lines obtained by chance and certain objects.

II. Evolution of the representation of the human figure.

1. First attempts at representation, similar to the preliminary stage.
2. The "tadpole" stage.
3. Transitional stage.
4. Complete representation of the human figure as seen in the full face.
5. Transitional stage between full face and profile.
6. The profile. (p. 16)

It was upon these premises that the use of human figure drawing tests became accepted measures of intelligence, especially the Goodenough-Harris Draw-a-Person test (G-H) (Harris, 1963). Children's drawings are developmental and change as the child learns new concepts. New concepts are incorporated into the intellect and are manifested in the drawing of a person. As the child develops, the conceptual aspects of the child's drawing become more profound and significant (Harris, 1963).

Unlike other standardized IQ tests, the validity and reliability of the G-H has not been easy to assess. Concurrent validity studies with other instruments consistently report only moderate correlations. Although disagreements arise as to its true usefulness, the fact remains that the G-H has been used not only in research, but in the assessment of children's intelligence over the past thirty years. Some psychologists use the G-H due to its ease of administration and enjoyment by young children (McCabe & Hilmo, 1985). McCabe and Hilmo (1985) presented a case in which use of the G-H proved to be more useful and a better predictor of a child's intellectual ability and progress than standard IQ measures due to its nonthreatening nature of administration.

Comparison of G-H with Other Measures of Intelligence

Tramill, Edwards and Tramill (1980) compared the G-H with the WISC-R using a sample of 100 children identified as experiencing academic difficulties. Both instruments were administered to all students by trained psychologists. Correlations between the total G-H score and each of the scores on the WISC-R were reported ($p<.01$). The highest correlations existed between the Full Scale IQ of the WISC-R and the Total G-H DAP IQ ($r=.63$).

Dunn (1967) conducted three separate studies comparing the G-H to other measures of intelligence. In the first study she administered the G-H and the Stanford-Binet L-M to 32 elementary children. In the second study he administered the G-H and the WISC to 93 public school children. In the third study he administered the G-H, the California Test of

Mental Maturity (CTMM), and the Iowa Test of Basic Skills to 90 elementary school children. Each study was conducted independently. Results of the first study reported a significant correlation of .78 between the Stanford-Binet L-M IQ and the G-H Draw a Man IQ ($p<.01$). The second study resulted in significant correlations between the G-H Draw a Man IQ and scores of the WISC ($p<.01$) (r ranged from .28 to .64). In the third study only one significant correlation ($p<.01$) was found: G-H Draw a Man IQ and CTMM Verbal IQ ($r=.32$). Although moderate correlations were reported between the G-H Draw a Man, the Stanford-Binet L-M and WISC, poor correlations were found with measures of achievement and G-H.

Olivier and Barclay (1967) compared the G-H with the Stanford-Binet L-M using a sample of 188 children enrolled in Head Start. The Stanford-Binet L-M was administered individually while the G-H was administered in groups. Moderate but significant correlations ($p<.01$) were found between G-H scores and scores on the Stanford-Binet L-M ranging from .40 to .65.

Comparison of the G-H with the PPVT and the Stanford-Binet L-M was assessed by Ritter, Duffey, and Fischman (1974). The sample of 31 kindergarten children were administered the three tests in random order. Moderate to low correlations were found between the Stanford-Binet L-M and G-H ($r=.55$), and the PPVT and G-H ($r=.37$).

White (1979) conducted a study comparing the G-H with the WISC-R, and Peabody Individual Achievement Test (PIAT).

Tests were administered to 30 public school children in a counterbalanced order. Again, moderate but significant ($p<.01$) correlations were found between the G-H scores and scores derived from the WISC-R (r ranged from .5 to .57) and PIAT (r ranged from .38 to .47).

Pihl and Nimrod (1976) conducted a study in which scores from the G-H collected from 44 children in the fifth grade were compared with teachers' ratings of academic achievement and with clinical psychologists "eyeball" assessments of the drawings. Significant ($p<.001$) but low correlations ($r=.30$) were found between teachers' judgements of children's academic ability and scores on the G-H. Moderate ($r=.46$ to .45) and significant correlations ($p<.001$) were found between psychologists "eyeball" assessment of the children's drawings and G-H scores.

Assessment of G-H Reliability and Validity

Test-retest reliability of the G-H was assessed in a study by McGilligan, Yater, and Huesing (1971). Forty-five subjects from a first grade classroom were administered the G-H in group sessions on two different occasions. The waiting period between testing was three weeks. There were no significant changes in the test scores. Computed reliabilities were in the moderate range (r ranged from .34 to .9) and significance at the .05 to .01 level was reached.

Dunn (1967) tested the inter- and intra-rater reliability of the G-H. The sample included 72 grade school children. Two independent raters scored the drawings after administration. One week later one of the original raters rescored the drawings. Inter-rater reliability was reported as $\underline{r}$=.88. Intra-rater reliability was reported as $\underline{r}$=.93. Both of these correlations are significant at the .01 level.

Strommen and Smith (1987) conducted research to test the internal consistency of the G-H. A sample of 150 children at four different age levels were administered the G-H in the study. KR-20's were used to assess internal consistency. The reliabilities were moderate to good ranging from .63 to .92. Reliabilities were highest for eight year olds and lowest for five year olds.

Scott (1981) conducted a review of the G-H and reported her findings of its validity and reliability. Her assessment agrees with the afore reviewed articles in that while reliability measures tend to be quite good, validity measures produce only moderate to low correlations. Although no conclusive evidence has been found to support the use of the G-H in determining special populations such as learning disabled, mentally retarded, or gifted, the G-H has been shown to be an adequate device for indicating below average intelligence (Scott, 1981; Tramill, Edwards, & Tramill, 1980).

Research with Naglieri's DAP

The DAP, constructed by Naglieri (1988), has yet to be extensively studied due to its relative newness and therefore has not yet established itself as a valid and reliable measure of intelligence. The DAP is based on the same theories as the G-H but was developed to answer criticisms of the G-H (Naglieri, 1988). It is postulated to have a more objective scoring system and more modern scoring criteria. The DAP also has new standardized norms. A few studies have been conducted, though, using this new instrument and will be reviewed here.

Prewett, Bardos, and Naglieri (1989) compared the DAP with the Matrix Analogies Test-Short Form (MAT-SF) and the Kaufman Test of Educational Achievement (KTEA). The sample consisted of 46 students identified as normal and 39 students identified as developmentally handicapped. The DAP and MAT-SF were administered in group settings while the KTEA was administered individually. For the normal group, the DAP correlated significantly (p<.05 to .01) with the MAT-SF. There was no significant relationship between the DAP and the KTEA for the normal sample. For the developmentally handicapped group correlations were low with only one reported as significant between the DAP and MAT-SF. Of the eight correlations reported for the DAP and KTEA in the developmentally handicapped group, five were significant at the .05 or .01 level.

Haddad and Juliano (1991) also compared the DAP with the MAT and with the Iowa Tests of Basic Skills (ITBS). The sample included 82 children in the fourth grade. The ITBS was administered to the children approximately one month before group administration of the DAP and MAT. The group correlation coefficient reported for the MAT and DAP was .32 ($p<.01$). The correlation coefficient reported for the DAP and ITBS was .47 ($p<.01$). These correlations, while significant, are in the low to moderate range.

Wisniewski and Naglieri (1989) compared the DAP with the WISC-R using a sample of 51 subjects referred for evaluation due to academic difficulties. The tests were administered in counterbalanced order. All correlations were significant ($p<.05$ to .01) and ranged from .31 to .54, which falls in the moderate category.

In a paper presented by Harrison, Schmitt, and Brown (1990) research was reviewed which compares the DAP, McCarthy Scales of Children's Abilities, and G-H. The sample consisted of 75 children from a kindergarten class. The children were administered the three tests in small groups. Eleven months later the children were administered the Stanford Achievement Test (SAT) and the Otis Lennon School Ability Test (OLSAT). The intercorrelation between the G-H and DAP was .86, which is statistically significant ($p<.001$). Correlations between the G-H and the achievement measures were low and nonsignificant. However, correlations between the DAP and the OLSAT and SAT were significant ($p<.05$). Thus it could be that the DAP has better predictive validity than the G-H.

Bardos, Softas, and Petrogiannis (1989) also compared the G-H and the DAP. Tests were group administered to a sample consisting of 114 students from Greece. All correlations were significant ($p<.001$) and ranged from .49 to .80.

Kamphaus and Pleiss (1991) reviewed aspects of the DAP, including validity and reliability. They essentially reviewed the previous articles already discussed and agree with this author. Reliability appears to be generally high, as reported by Naglieri (1988), with an internal consistency coefficient of .86. Test-retest reliability is reported between .60 and .89. Validity of the measure is only reported as moderate or mediocre as yet.

Summary and Conclusions

In reviewing the literature it was found that the SB:FE, DAP, and G-H have not yet been compared. In order to support the theory that the DAP is an adequate measure of intelligence, as defined by the SB:FE, some test relationships must be discussed. In the literature review it was found that the SB:FE correlated significantly with the Stanford-Binet L-M (Hartwig, Sap, & Clayton, 1987; Lukens, 1988; Clark, Wortman, Warnock, & Swerdlik, 1987). The G-H was also found to correlate significantly with the Stanford-Binet L-M (Dunn, 1967; Olivier & Barclay, 1967). It could thus be proposed that the SB:FE and the G-H would correlate significantly if compared.

Also, the SB:FE was found to correlate significantly with the WISC-R (Tramill, Edwards, & Tramill, 1980; Greene, Sapp, & Chissom, 1990; Phelps, 1989; Hollinger & Baldwin, 1990; Lukens, 1990; Phelps, Bell, & Scott, 1988; &

Rothlisberg, 1987). The G-H was also found to correlate significantly with the WISC-R (White, 1979). The DAP was also found to correlate significantly with the WISC-R (Wisniewski & Naglieri, 1989) Thus, it could be proposed that the SB:FE would correlate significantly with both the G-H and DAP if compared. Two articles reviewed reported that G-H and DAP are similar measures and when compared produce significant correlations (Bardos, Softas, & Petrogiannis, 1989; Harrison, Schmitt, & Brown, 1990).

In summary, after an extensive look at the articles which review the DAP, G-H, and SB:FE, it appears that evidence does exist to predict that, upon comparison, significant relationships will exist among these three measures.

Research Hypotheses

Based on the literature reviewed, the following hypotheses will be tested:

1. The first set of hypotheses will examine the relationship between the DAP and SB:FE area scores (i.e., Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-term Memory) and the SB:FE composite Score.

(a) The DAP man score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled.

(b) The DAP woman score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled.

(c) The DAP self score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled.

(d) The DAP composite score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled.

(e) The DAP composite score will show a significant relationship with the SB:FE composite score.

2. The second set of hypotheses will examine the relationship between the G-H and SB:FE area scores (i.e., Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory) and the SB:FE composite score.

(a) The G-H man score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled.

(b) The G-H woman score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled.

(c) The G-H composite score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled.

(d) The G-H composite score will show a significant relationship with the SB:FE composite score.

3. The magnitude of the correlations between DAP composite and SB:FE composite will be significantly greater than the magnitude of the correlation between G-H Composite and SB:FE Composite.

CHAPTER III

METHODOLOGY

The data used in this project was originally collected for another study entitled "Children's Picture Drawing, Cognitive Functioning and Neuromotor Development" (Tomes & Heilbuth, 1991).

Subjects

The total sample consisted of 72 children which were divided into four groups: 18 five-year-old boys (*M* = 5 yr 5 mo; range is 5-3 to 5-11), 18 five-year-old girls (*M* = 5 yr 4 mo; range is 5-2 to 5-11), 21 seven-year-old boys (*M* = 7 yr 3 mo; range is 7-0 to 7-11), and 15 seven-year-old girls (*M* = 7 yr 4 mo; range is 7-1 to 7-11). The sample population consisted of predominantly white children with varying socioeconomic backgrounds ranging from low- to upper- middle class. The seven-year-old subjects were recruited from two public elementary schools. Thirty-three of the five-year-old subjects attended half-day kindergarten programs and the remaining thirteen were from a half-day preschool program. Participation was on a voluntary basis. For purposes of the present study, only data from the seven-year-old children who participated in the original study

(Tomes & Heilbuth, 1991) were used.

Procedures

Subjects were recruited through a public school system and through a private preschool program in a small midwestern town. An explanation of the study and parental consent forms were sent home with children from school after obtaining permission from the school superintendent and principal. Parents were asked to sign the forms and return them to the school. Only those children whose parents returned signed consent forms were included in the study. Subject confidentiality was ensured through a numbering system and maintained throughout the original study as well as this project.

After obtaining consent forms, each child was tested in three separate sessions at the school during regular school hours. Examiners attempted to accommodate each child's school schedule. In the first session, which took approximately forty-five minutes to an hour, children were administered three tests, one of which was the G-H. In the second session the child was administered the Stanford-Binet: Fourth Edition, which took approximately one hour to 90 minutes. During the administration of the SB:FE, breaks were taken as needed according to each child's behavior (i.e. restless, tired, yawning). In the third session each child was administered two motor assessment scales. Data were collected by three trained graduate students including the author of this thesis. Each examiner was responsible for giving a specific test or group of tests to each child involved in the study during the course of the research.

All tests were scored according to standard protocols described in their respective test manuals. For purposes of the present research the G-H drawings were rescored using the DAP scoring criteria, resulting in man, woman, self, and composite DAP scores.

Instruments

This study compared three different measures of intelligence: (1) The Stanford-Binet Intelligence Scale: Fourth Edition (Thorndike, Hagen, & Sattler, 1986), (2) Draw A Person: A Quantitative Scoring System (Naglieri, 1988), and (3) The Goodenough-Harris Drawing Test (Harris, 1963).

Stanford-Binet: Fourth Edition

The Stanford-Binet: Fourth Edition (SB:FE) provides a range of tasks used to assess intelligence from age two to adult. Revised in 1986, the SB:FE provides an overall intelligence score which comprises four area scores: Verbal Reasoning, Quantitative Reasoning, Abstract/Visual Reasoning, and Short-Term Memory. The SB:FE was based on a three level hierarchical model of the structure of cognitive abilities (see Figure 1). At the top of this model lies g, the general reasoning factor. The g is broken into three abilities: crystallized abilities, fluid-analytic abilities, and short-term memory. At the third level, under crystallized abilities, lies verbal reasoning and quantitative reasoning. Under fluid analytic abilities lies abstract/visual reasoning.

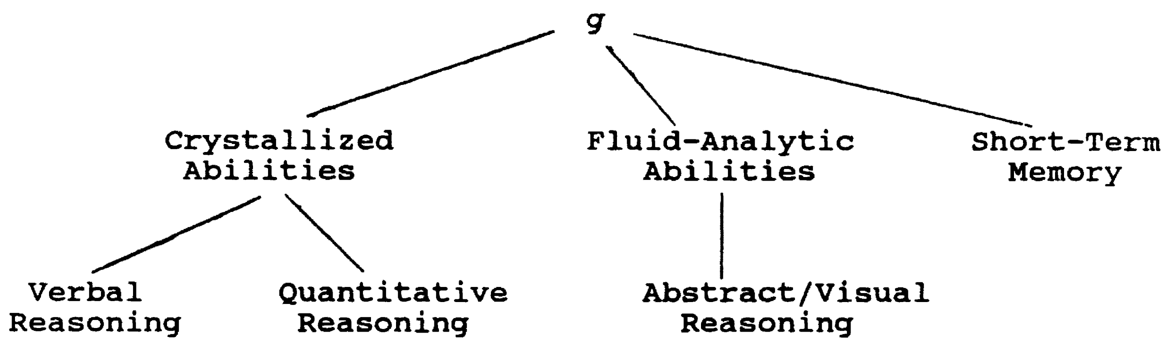

Figure 1. Three level hierarchical model for SB:FE

The SB:FE contains 15 tests: four for the Verbal Reasoning Area, four for the Abstract/Visual Area, three for the Quantitative Reasoning Area, and four for the Short-Term Memory Area. The SB:FE yields four area standard scores and a composite IQ standard score. Each of these five scores is based on a mean of 100 and a standard deviation of 16.

Selection of tests to be administered is made based on age and highest item passed in Vocabulary, the first test administered. Each item is scored on a pass/fail basis. There are no time limits for completion of tests in the SB:FE though time for some items are recorded by the examiner. Therefore, administration takes approximately one hour to 90 minutes. Subjects are thus given breaks if they become tired or restless during testing.

According to the SB:FE Technical Manual (Thorndike, Hagen & Sattler, 1986) KR-20 reliabilities were calculated for each of the four areas. Results are as follows: The Verbal Reasoning Area KR-20's range from .74 to .96; Abstract/Visual Area KR-20's range from .74 to .96; Quantitative Reasoning Area KR-20's range from .80 to .95; and the Short-Term Memory Area KR-20's range from .66 to .95. The overall Composite Standard Area Score reliability ranges from .95 to .99 with a SEM of 3.6 to 1.6. Test-retest reliabilities were also reported in the Test Manual. Test-retest reliability of the Composite SAS in a preschool sample was .91 and in an elementary sample was .90.

Validity was assessed using three kinds of studies which are detailed in the Test Manual. Factor analysis was done to determine construct validity of the SB:FE. Results indicate that all 15 tests had substantial loadings on *g*. Confirmation of the four area scores was also received. Criterion validity was assessed by comparing the SB:FE with each of the following: Stanford-Binet Intelligence Scale: Form L-M, the Wechsler Intelligence Scale for Children-Revised (WISC-R), the Wechsler Adult Intelligence Scale-Revised (WAIS-R), the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), and the Kaufman Assessment Battery for Children (K-ABC). Correlations between each of these tests and the SB:FE were high, ranging from .8 to .91 on composite and full scale scores. A series of studies was also done in which the SB:FE was administered to special groups of participants. Results indicate that the SB:FE can reliably discriminate between mentally retarded and learning disabled examinees. Others have reviewed the validity of the SB:FE and found it to be "... at least as good a measure of *g* as other currently available tests of intellectual ability" (Laurent, Swerdlik, & Ryburn, 1986, p. 110). Thus, the SB:FE appears adequate as a comparison test for the DAP and Goodenough-Harris.

Draw a Person: A Quantitative Scoring System

The DAP was developed in 1988 to provide a nonverbal measure of ability which can be administered quickly, and scored objectively. The test requires children to first draw a man, then a women, and lastly a picture of themselves. Testing takes approximately fifteen minutes. The examinee is given five minutes to complete each drawing. The test renders a composite standard score which is composed of the individual scores of the man, woman, and self drawings, as well as standard scores for each of the individual drawings. These four scores are standard scores based on a mean of 100 and a standard deviation of 15.

According to the DAP manual (Naglieri, 1988), reliability was assessed by computing coefficients for internal consistency, test-retest reliability, and interrater reliability. Cronbach's alphas were calculated to assess internal consistency. Across ages five to seventeen the alpha ranged from .83 to .89 for the composite score. When the scores are broken down into man, woman, and self the alphas range from .56 to .78. Therefore internal consistency for the total score is higher than for individual scores. The test-retest reliability for the composite score ranged from .60 to .89 while coefficients for man, woman, and self scores ranged from .21 to .92. The interrater reliability coefficients for man, woman, self, and composite are .92, .93, .93, .95, respectively.

Both construct and criterion related validity were reported in the DAP manual (Naglieri, 1988). Construct validity was assessed by examining the developmental change in mean scores. The mean raw scores showed an increase with age of examinee. Also, a significant correlation ($p<.01$) was found between age and man, woman, self, and composite scores for examinees aged five to eleven. Criterion related validity was assessed by comparing the DAP with the Matrix Analogies Test-Short Form (MAT-SF) and the Multilevel Academic Survey Test (MAST). The MAT-SF is a measure of nonverbal ability, as is the DAP. The MAST is an achievement test consisting of Reading and Mathematics sections which were used in the correlation with the DAP. All DAP scores correlated significantly ($p<.01$) with both the MAT-SF and the MAST although DAP scores correlated higher with the MAT.SF.

Goodenough-Harris Drawing Test

The G-H is also a test designed to be a measure of intellectual development. The administration techniques are similar to those of the DAP. The examinee is asked to first draw a picture of a man. Once that picture is completed the examinee is asked to draw a woman. Lastly the examinee is asked to draw a picture of him- or herself. Scoring procedures are available for the man and woman drawings but not for the self drawing. Standard scores rendered from the test include man, woman, and total scores, which are based on a mean of 100 and a standard deviation of 15. A scale is also available to measure the quality of the individual drawings (man, woman, and self), although these scores are not used in the present research.

Reliability of the G-H was assessed by computing the intercorrelations between two independent scores. For the man and woman scale the intercorrelations ranged from .91 to .98 (Harris, 1963). Validity of the test was assessed by correlating the G-H with other measures, including the Stanford-Binet: Third Edition, and the Wechsler Intelligence Scale for Children. Correlations ranged from .26 to .72.

Data Analysis

Pearson product-moment correlations were calculated to determine significant relationships between variables listed in each hypothesis. Multivariate multiple regression analyses were used to test to what extent each drawing score predicts SB:FE area scores as a set. It was planned to use Roy Bargman stepdown _F_ tests for significant multivariate results. Since no multivariate multiple regressions were significant, univariate _F_-tests were calculated on each drawing score and the individual SB:FE area scores to test hypotheses #1 and #2. Simple regression analyses were used to test hypotheses #1e and #2d. Finally, the Fisher's Z transformation was used to test hypothesis #3.

CHAPTER IV

RESULTS OF THE STUDY

This chapter reports results of data analysis from 36 seven-year-old children. First, means and standard deviations will be reported. Secondly, a description of the Pearson Product Moment Correlations obtained will be presented. Third, results pertaining to each hypothesis will be reported.

Findings

Means and standard deviations from each of the variables in the study are reported in Table I. Scores from the SB:FE were consistently higher than scores from the DAP or G-H. Means reported in the SB:FE Technical Manual (Thorndike et al, 1986) for seven year olds range from $\underline{M}$=100.2 to 101.5 with standard deviations from 13.1 to 16.0. Means reported for the current sample fall below that range, with the exception of Short-Term Memory which is higher ($\underline{M}$=102.6). Standard deviations, as well, fall below those reported in the Technical Manual with the exception of Quantitative Reasoning and Short-Term Memory ($\underline{SD}$=14.2 to 14.4).

TABLE I

MEANS AND STANDARD DEVIATIONS FOR THE
SB:FE, DAP, AND G-H

Scale n=36	*M*	*SD*
SB:FE Composite	97.5	10.0
SB:FE Verbal Reasoning	98.4	8.7
SB:FE Abstract/Visual Reasoning	92.4	12.0
SB:FE Quantitative Reasoning	98.1	14.2
SB:FE Short-Term Memory	102.7	14.4
G-H man	97.0	15.8
G-H woman	91.4	13.8
G-H composite	94.4	14.1
DAP man	87.7	17.0
DAP woman	91.1	15.9
DAP self	88.6	14.7
DAP composite	87.3	16.5

Point score means and standard deviations rather than standard score means and standard deviations are reported in the G-H manual (Harris, 1963) so comparisons cannot be made. Means from the G-H are higher than the DAP means. The DAP manual reports similar findings: the DAP mean ranged from 88.3 to 91.9 and the G-H means ranged from 94.2 to 99.8 (Naglieri, 1988). The DAP man and woman standard deviations in the current study were higher than those reported in the manual while DAP self standard deviations were very similar ($\underline{SD}$=14.6 vs 14.7).

Pearson Product Moment Correlations were computed across all scores and the results are reported in Tables II through IV. Intra-test correlations for the SB:FE were significant ($\underline{p}<.05$) with the exception of SB:FE Quantitative Reasoning with each of the other three area scores. Highly significant inter-test correlations were found for both the DAP and G-H ($\underline{p}<.01$). Correlations of DAP with G-H scores were also significant ($\underline{p}<.01$). These results are reported in Table III.

TABLE II

CORRELATION COEFFICIENTS FOR SB:FE

	SBIQ[a]	SBVR[b]	SBAVR[c]	SBQR[d]	SBSTM[e]
SBIQ[a]	1.00	.553**	.685**	.621**	.801**
SBVR[b]		1.00	.350*	-.011	.391*
SBAVR[c]			1.00	.181	.374*
SBQR[d]				1.00	.312
SBSTM[e]					1.00

[a] SB:FE Composite
[b] SB:FE Verbal Reasoning
[c] SB:FE Abstract/Visual Reasoning
[d] SB:FE Quantitative Reasoning
[e] SB:FE Short-Term Memory

*$\underline{p}<.05$; **$\underline{p}<.01$

Correlation coefficients of SB:FE with DAP and G-H are reported in Table IV. Only three correlations are significant between the drawing tests and SB:FE: G-H woman with SB:FE Abstract/Visual Reasoning (r=.38, p<.05), DAP self with SB:FE Composite (r=.34, p<.05), and DAP self with SB:FE Abstract/Visual Reasoning (r=.38, p<.05).

TABLE III

CORRELATION COEFFICIENTS FOR DRAWING TESTS

	GHMAN	GHWOM	GHAVG	DAPMAN	DAPWOM	DAPSF	DAPTOT
GHMAN[a]	1.00	.816**	.959**	.923**	.782**	.644**	.877**
GHWOM[b]		1.00	.947**	.767**	.872**	.680**	.863**
GHAVG[c]			1.00	.892**	.864**	.690**	.913**
DAPMAN[d]				1.00	.816**	.586**	.903**
DAPWOM[e]					1.00	.716**	.945**
DAPSF[f]						1.00	.841**
DAPTOT[g]							1.00

[a] G-H man score
[b] G-H woman score
[c] G-H composite score
[d] DAP man score
[e] DAP woman score
[f] DAP self score
[g] DAP composite score

** p<.01

TABLE IV

CORRELATION COEFFICIENTS FOR SB:FE, DAP, AND G-H

	SBIQ[h]	SBVR[i]	SBAVR[j]	SBQR[k]	SBSTM[l]
GHMAN[a]	.234	.093	.195	.026	.287
GHWOM[b]	.290	-.034	.376*	.064	.312
GHAVG[c]	.269	.029	.290	.048	.309
DAPMAN[d]	.224	.140	.209	-.002	.253
DAPWOM[e]	.142	-.095	.258	-.084	.248
DAPSF[f]	.335*	.154	.379*	.045	.322
DAPTOT[g]	.261	.074	.310	-.014	.310

[a] G-H man score
[b] G-H woman score
[c] G-H composite score
[d] DAP man score
[e] DAP woman score
[f] DAP self score
[g] DAP composite score
[h] SB:FE Composite
[i] SB:FE Verbal Reasoning
[j] SB:FE Abstract/Visual Reasoning
[k] SB:FE Quantitative Reasoning
[l] SB:FE Short-Term Memory

* $p<.05$

Hypothesis #1a.

The DAP man score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled. The multivariate test of the relationship between the DAP man score and the SB:FE area scores taken as a set was not significant. With 4 and 31 degrees of freedom, the Wilks Lamda result was .91, interpreted as an exact *F* of .75. The probability of obtaining an *F*-value of this magnitude is .567. Because the desired .05 significance level was not reached the stepdown procedures were not used. Univariate *F*-tests with 1 and 34 degrees of freedom are reported in Table V. No support was found for Hypothesis #1a.

TABLE V

UNIVARIATE REGRESSION ANALYSIS OF DAPMAN ON SB:FE AREA SCORES

Variable	*r*	Adj. r^2	*F*-value	Sign. of *F*
SB:FE Short-Term Memory	.25	.06	2.32	.137
SB:FE Abstract/Visual	.21	.04	1.55	.221
SB:FE Verbal	.14	.02	.68	.416
SB:FE Quantitative	.00	.00	.00	.991

Hypothesis #1b.

The DAP woman score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled. The multivariate test of the relationship between the DAP woman score and SB:FE area scores taken as a set was not significant. With 4 and 31 degrees of freedom, the Wilks Lamda result was .79, interpreted as an exact *F* of 2.12. The probability of obtaining an *F*-value of this magnitude is .102. Because the desired .05 significance level was not reached stepdown procedures were not used. Univariate *F*-tests with 1 and 34 degrees of freedom are reported in Table VI. Significance was not reached and therefore no support was found for Hypothesis #1b.

TABLE VI

UNIVARIATE REGRESSION ANALYSIS OF DAPWOMAN ON SB:FE AREA SCORES

Variable	*r*	Adj. r^2	*F*-value	Sign. of *F*
SB:FE Short-Term Memory	.25	.06	2.23	.144
SB:FE Abstract/Visual	.26	.07	2.43	.128
SB:FE Verbal	.10	.01	.31	.580
SB:FE Quantitative	.08	.01	.24	.625

Hypothesis #1c.

The DAP self score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled. The multivariate test of the relationship between the DAP self score and the SB:FE area scores as a set was not significant. With 4 and 31 degrees of freedom, the Wilks Lamda result was .81, interpreted as an exact *F* of 1.82. The probability of obtaining an *F*-value of that magnitude is .151. Because the desired .05 significance level was not achieved stepdown procedures were not used. Univariate *F*-tests with 1 and 34 degrees of freedom are reported in Table VII. One *F*-value reached significance ($p<.03$). The DAP self score predicts SB:FE Abstract/Visual Reasoning ($F(1,34)=5.68$, $p<.05$). The DAP self score accounted for 14% of the variance in the Abstract/Visual Reasoning area. Also, as mentioned earlier a significant correlation was found between DAP self score and SB:FE Abstract/Visual Reasoning ($r=.379$, $p<.05$) and DAP self score and SB:FE Composite ($r=.335$, $p<.05$). Therefore it appears that some support was found for Hypothesis #1c.

TABLE VII

UNIVARIATE REGRESSION ANALYSIS OF DAPSELF ON SB:FE AREA SCORES

Variable	r	Adj. r^2	F-value	Sign. of F
SB:FE Short-Term Memory	.32	.10	3.93	.055
SB:FE Abstract/Visual	.38	.14	5.68	.023
SB:FE Verbal	.15	.02	.83	.369
SB:FE Quantitative	.05	.00	.07	.794

Hypothesis #1d.

The DAP composite score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled. The multivariate test of the relationship between the DAP composite score and SB:FE area scores taken as a set was not significant. With 4 and 31 degrees of freedom, the Wilks Lamda result was .83, interpreted as an exact F of 1.63. The probability of obtaining an F-value of that magnitude is .192. Because the desired .05 significance level was not reached stepdown procedures were not used. Univariate F-tests with 1 and 34 degrees of freedom are reported in Table VIII. Significance was not reached across any of the SB:FE area scores, therefore no support was found for Hypothesis #1d.

TABLE VIII

UNIVARIATE REGRESSION ANALYSIS OF DAP COMPOSITE ON SB:FE AREA SCORES

Variable	*r*	Adj. r^2	*F*-Value	Sign. of *F*
SB:FE Short-Term Memory	.31	.10	3.61	.066
SB:FE Abstract/Visual	.31	.10	3.61	.066
SB:FE Verbal	.07	.01	.19	.667
SB:FE Quantitative	.01	.00	.01	.936

Hypothesis #1e.

The DAP composite score will show a significant relationship with the SB:FE composite score. The relationship between DAP composite and SB:FE composite score as tested by simple regression was not significant. With 1 and 34 degrees of freedom an *F* value of 2.49 was attained. The probability of attaining an *F* value of this magnitude is .124. As noted earlier a significant correlation was not reached between the DAP composite and SB:FE composite. Because a significant relationship was not found between these two variables no support was rendered for Hypothesis #1e.

Hypothesis #2a.

The G-H man score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled. The multivariate test of the relationship between the G-H man score and the SB:FE area scores taken as a set was not significant. With 4 and 31 degrees of freedom, the Wilks Lamda result was .90, interpreted as an exact $\underline{F}$ of .86. The probability of obtaining an $\underline{F}$-value of that magnitude is .497. Because the desired .05 significance was not reached stepdown procedures were not used. Univariate $\underline{F}$-tests with 1 and 34 degrees of freedom are reported in Table IX. No significant relationships were found therefore support was not obtained for Hypothesis #2a.

TABLE IX

UNIVARIATE REGRESSION ANALYSIS OF G-H MAN ON SB:FE AREA SCORES

Variable	$\underline{r}$	Adj. $\underline{r}^2$	$\underline{F}$-value	Sign. of $\underline{F}$
SB:FE Short-Term Memory	.29	.08	3.05	.090
SB:FE Abstract/Visual	.20	.04	1.35	.254
SB:FE Verbal	.09	.01	.30	.589
SB:FE Quantitative	.03	.00	.02	.880

<u>Hypothesis #2b</u>.

The G-H woman score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled. The multivariate test of the relationship between the G-H woman score and the SB:FE area scores taken as a set was not significant. With 4 and 31 degrees of freedom, the Wilks Lamda result was .76, interpreted as an exact *F* of 2.50. The probability of obtaining an *F*-value of this magnitude is .063. Because the desired .05 significance level was not reached stepdown procedures were not used. Univariate *F*-tests with 1 and 34 degrees of freedom are reported in Table X. A significant relationship was found between SB:FE Abstract/Visual Reasoning and G-H woman scores ($F(1,34)=5.58$, $p<.03$). It was also reported earlier that the G-H woman score correlated significantly with SB:FE Abstract/Visual Reasoning score ($r=.376$, $p<.05$). Therefore, support was found for Hypothesis #2b.

TABLE X

UNIVARIATE REGRESSION ANALYSIS OF G-H WOMAN ON SB:FE AREA SCORES

Variable	r	Adj. r^2	F-Value	Sign. of F
SB:FE Short-Term Memory	.31	.10	3.66	.064
SB:FE Abstract/Visual	.38	.14	5.58	.024
SB:FE Verbal	.03	.00	.04	.846
SB:FE Quantitative	.06	.00	.14	.712

Hypothesis #2c.

The G-H composite score will show a significant relationship with the SB:FE area scores when the influence of the others is controlled. The multivariate test of the relationship between the G-H composite score and the SB:FE area scores taken as a set was not significant. With 4 and 31 degrees of freedom, the Wilks Lamda result was .84, interpreted as an exact *F* of 1.50. The probability of obtaining an *F*-value of this magnitude is .227. Because the desired .05 significance level was not reached stepdown procedures were not used. Univariate *F*-tests with 1 and 34 degrees of freedom are reported in Table XI. No significant *F*-values were found, therefore no support was obtained for Hypothesis #2c.

TABLE XI

UNIVARIATE REGRESSION ANALYSIS OF G-H COMPOSITE ON SB:FE AREA SCORES

Variable	*r*	Adj. r^2	*F*-Value	Sign. of *F*
SB:FE Short-Term Memory	.31	.10	3.59	.067
SB:FE Abstract/Visual	.29	.08	3.13	.086
SB:FE Verbal	.03	.00	.03	.867
SB:FE Quantitative	.05	.00	.08	.783

Hypothesis #2d.

The G-H composite score will show a significant relationship with the SB:FE composite score. The simple regression test of the relationship between G-H composite and SB:FE composite score was not significant. With 1 and 34 degrees of freedom an $\underline{F}$-value of 2.66 was attained. The probability of obtaining an $\underline{F}$-value of this magnitude is .112. As reported above, the correlation between G-H composite and SB:FE composite scores also did not reach significance. No support was found for Hypothesis #2d.

Hypothesis #3.

The magnitude of the correlations between DAP composite and SB:FE composite will be significantly greater than the magnitude of the correlation between G-H composite and SB:FE composite. To test this hypothesis, the Fisher's $\underline{Z}$ transformation was used. This statistical method was used to determine if there was a significant difference between the SB:FE composite with DAP composite correlation ($\underline{r}$=.26) and the SB:FE composite with G-H composite correlation ($\underline{r}$=.27). Because these correlations are so similar is was not suprising that the $\underline{Z}$ value was only .15, which does not reach the desired .05 significance. Therefore no support was found for Hypothesis #3.

Summary

Out of the ten hypotheses predicted, only two found some support. The DAP self score was found to predict 14% of the varience of the SB:FE Abstract/Visual Reasoning area ($F(1,34)=5.68$, $p<.05$). Also, a significant correlation was found between DAP self and SB:FE Abstract/Visual Reasoning ($r=.379$, $p<.05$). G-H woman score was found to predict 14% of the varience of SB:FE Abstract/Visual Reasoning score ($F(1,34)=5.58$, $p<.03$). A significant correlation was also reported between SB:FE Abstract/Visual Reasoning and G-H woman score ($r=.376$, $p<.05$).

CHAPTER V

DISCUSSION

Summary

The purpose of this study was to investigate relationships between the SB:FE, DAP, and G-H. Because two of these three measures are relatively new (SB:FE and DAP), little research has been done which investigates their validity. Hypotheses for this study were based on the current literature available which indicated that the DAP and G-H would have significant relationships with the newly revised SB:FE. Surprisingly though, results from the current research do not significantly support the majority of hypotheses. Using the multivariate multiple regression analysis no significant F-values were reported, therefore univariate F-tests were reported rather than stepdown F-tests. Because the drawing tests appeared to only tap two of the four area scores in the SB:FE it is not surprising that the drawing tests did not predict the area scores as a set.

To understand these results it is important to bring up an original question in this study. Do drawing tests assess the same constructs as the SB:FE? If we review Table 3, we find that while correlations, for the most part, were not significant, higher Pearson r's were found between each drawing test score and two area scores of the SB:FE (Short-Term Memory and Abstract/Visual Reasoning). In fact, two of the three significant correlations fall in the SB:FE Abstract/Visual Reasoning area. In the area of Abstract/Visual Reasoning, Pearson r's range from .20 to .38 across both sets of drawing scores. In the Short-Term Memory area Pearson r's range from .25 to .32 across both sets of drawing scores. Verbal Reasoning and Quantitative Reasoning Pearson r's are all less than .15. The SB:FE Composite score is made up of all area scores so it is not surprising that these Pearson r's fall in the middle (r=.14 to .34). Analyzing the data in this way makes it possible to answer the question posed above. It appears that both the DAP and G-H are tapping into two specific parts (or factors) of the whole which makes up intelligence as defined by the SB:FE. We can thus propose that standardized intelligence scales such as the SB:FE and drawing tests do use similar constructs but drawing tests only tap into a smaller part of that construct, primarily the child's ability to use abstract-visual thinking and his/her short term memory capacity or capability.

Data from 36 five-year-old children was also analyzed and the same patterns arose in the correlation coefficients. But, significance was reached across the Abstract/Visual Reasoning area and all drawing test scores ($p<.05$). The five-year-old's data analysis rendered higher Pearson r's across all correlations suggesting a possibility of significant age related differences between the SB:FE, DAP and G-H. This is a possible problem that needs to be addressed in later research.

It is important to note that the correlations between the G-H and DAP scores were highly significant ($p<.01$). After calculating the Fisher's Z ($Z=.15$), it can be said that the two tests are very similar in their correlations to the SB:FE. There was no significant difference in the correlations between G-H composite with SB:FE composite and DAP composite with SB:FE composite. This suggests that the DAP and G-H are testing the same construct.

It is also interesting to note the two significant correlations that include the DAP self score. The DAP self is the only drawing score that significantly correlated with the SB:FE Composite score ($r=.335$, $p<.05$). The DAP self was also the only significant predictor of the SB:FE Composite ($F=4.3$, $p<.05$), although no hypothesis was raised to test this relationship. The DAP self score also rendered a significant univariate F-value of 5.7 ($p<.03$) as a predictor of SB:FE Abstract/Visual Reasoning. One explanation of this result could be that children perform better on the third drawing. If children perform better on the third drawing, due to being more relaxed and possibly striving to do better on a "self portrait", then they will have a higher self score which would correlate more highly with the SB:FE score. Ten of the drawings did have better DAP self scores than man, woman, or composite scores.

The relationship between the SB:FE Abstract/Visual Reasoning and the G-H woman score also rendered a significant correlation (r=.38, p<.05). The G-H woman score was also a significant predictor of the SB:FE Abstract/Visual Reasoning score with an F-value of 5.58 (p<.03). Scoring procedures for the man and woman drawings are different in the G-H and therefore may account for the differences reported here.

No other significant relationships were found after analyzing the data. As mentioned before, analysis using data from a five-year-old sample resulted in nine more significant correlations than were found for the seven-year-old sample. Therefore, it is possible to conclude that the child's age plays an important part in the assessment of relationships between SB:FE and the drawing tests. In reviewing the literature, it was found that a majority of the studies which reported results using the DAP or G-H collapsed across ages in their data analysis which could account for their higher correlations.

Conclusions

Although the results of this study are not as predicted, the research in itself has not lost it's value. Important conclusions can still be drawn from it as well as implications for future research. After analyzing the results one can conclude that the DAP and G-H are similar tools. One appears no more valuable than the other with the exception of gaining a self score using the DAP, which does appear to correlate and predict SB:FE Abstract/Visual Reasoning scores better than any of the other drawing scores from either test. The addition of the self score appears to be the only improvement the DAP makes over the G-H. Overall, neither drawing test gives a more comprehensive assessment of intelligence than the SB:FE. The DAP and G-H appear to only "scratch the surface" when it comes to predicting and assessing intelligence. They may be most useful as part of a battery of tests used to assess a child's overall development. Therefore, they are not recommended to be used in place of the SB:FE.

While the DAP may not be a satisfactory predictor of the SB:FE, it must not be assumed that it is of no value. To consider the SB:FE the "only" way to measure intelligence or even the best way would be reckless. The SB:FE is itself still under scrutiny (Laurent, Swerdlik, & Ryburn, 1992; Gridley & McIntosh, 1991; Ownby & Carmin, 1988; Thorndike, 1990; and Kline, 1989). It must be remembered that intelligence is still defined by different people several different ways and to suggest that one measure of any form gives a complete picture of any individual child's intelligence would be inaccurate.

Because this study is the first currently known by the author to compare these three measures it is recommended that other studies be done to replicate the findings and thus provide stronger validity and reliability for this study's findings. It is also suggested that in the future data collection and analysis of these measures be conducted across a variety of ages and comparisons be made (specifically ages 5-8). A comparison of the DAP with other measures of intelligence, and possibly achievement (Harrison, Schmitt, & Brown, 1990), is also suggested to determine if it correlates or predicts measures other than the SB:FE significantly better.

REFERENCES

Bardos, A.N., Softas, B.C., & Petrogiannis, K. (1989). Comparison of the Goodenough-Harris and Naglieri's Draw-a-Person scoring systems for Greek children. School Psychology International, 10, 205-209.

Carvajal, H. (1988). Relationship between scores of gifted children on Stanford-Binet IV and Peabody Picture Vocabulary Test - Revised. Diagnostique, 14(1), 22-25.

Carvajal, H., Hardy, K., Smith, K.L., & Weaver, K.A. (1988). Relationships between scores on Stanford-Binet IV and Wechsler Preschool and Primary Scale of Intelligence. Psychology in the Schools, 25, 129-131.

Carvajal, H.H., Parks, J.P., Bays, K.J., Logan, R.A., Lujano, C.I., Page, G.L., & Weaver, K.A. (1991). Relationships between scores on Wechsler Preschool and Primary Scale of Intelligence - Revised and Stanford-Binet IV. Psychological Reports, 69, 23-26.

Carvajal, H., & Weaver, K.A. (1989). Relationships between scores of gifted children on Stanford-Binet IV and Wechsler Intelligence Scale for Children-Revised. Diagnostique, 14(2), 89-93.

Clark, R.D., Wortman, S., Warnock, S., & Swerdlik, M. (1987). A correlational study of form L-M and the 4th edition of the Stanford-Binet with 3- to 6-year olds. Diagnostique, 12(2), 118-120.

Dunn, J.A. (1967). Inter- and intra-rater reliability of the new Harris-Goodenough Draw-a-Man test. Perceptual and Motor Skills, 24, 269-270.

Dunn, J.A. (1967). Validity coefficients for the new Harris-Goodenough Draw-a-Man test. Perceptual and Motor Skills, 24, 299-301.

Goodenough, F.L. (1926). Measurement of intelligence by drawings. New York: World Book.

Greene, A.C., Sapp, G.L., & Chissom, B. (1990). Validation of the Stanford-Binet Intelligence Scale: Fourth Edition with exceptional black male students. Psychology in the Schools, 27, 35-41.

Gridley, B.E., & McIntosh, D.E. (1991). Confirmatory factor analysis of the Stanford-Binet: Fourth Edition for a normal sample. Journal of School Psychology, 29, 237-248.

Haddad, F.A., & Juliano, J.M. (1991). Relations among scores on Matrix Analogies Test, Draw-a-Person, and the Iowa Tests of Basic Skills for low socioeconomic children. Psychological Reports, 69, 299-302.

Harris, D.B. (1963). Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace, and World.

Harrison, P.L., Schmitt, C.S., & Brown, L.H. (1990). Comparison of three draw a person scoring systems for young children. Paper presented at the meeting of the National Association of School Psychologists, San Francisco, CA.

Hartwig, S.S., Sapp, G.L., & Clayton, G.A. (1987). Comparison of the Stanford-Binet Intelligence Scale: form L-M and the Stanford-Binet Intelligence Scale Fourth Edition. Psychological Reports, 60, 1215-1218.

Hayden, D.C., Furlong, M.J., & Linnemeyer, S. (1988). A comparison of the Kaufman Assessment Battery for Children and the Stanford-Binet IV for the assessment of gifted children. Psychology in the Schools, 25, 239-243.

Hendershott, J.L., Searight, H.R., Hatfield, J.L., & Rogers, B.J. (1990). Correlations between the Stanford-Binet, Fourth Edition and the Kaufman Assessment Battery for Children for a preschool sample. Perceptual and Motor Skills, 71, 819-825.

Hollinger, C.L., & Baldwin, C. (1990). Comparing scores on the Stanford-Binet, Fourth Edition with the WISC-R for exceptional children. Psychological Reports, 66, 979-984.

Kamphaus, R.W., & Pleiss, K.L. (1991). Draw-a-person techniques: Tests in search of a construct. Journal of School Psychology, 29, 395-401.

Kline, R.B. (1989). Is the Fourth Edition Stanford-Binet a four-factor test? Confirmatory factor analyses of alternative models for ages 2 through 23. Journal of Psychoeducational Assessment, 7, 4-13.

Knight, B.C., Baker, E.H., & Minder, C.C. (1990). Concurrent validity of the Stanford-Binet: Fourth Edition and Kaufman Assessment Battery for Children with learning-disabled students. Psychology in the Schools, 27, 116-120.

Krohn, E.J., & Lamp, R.E. (1989). Concurrent validity of the Stanford-Binet Fourth Edition and K-ABC for head start children. Journal of School Psychology, 27, 59-67.

Lamp, R.E., & Krohn, E.J. (1990). Stability of the Stanford-Binet Fourth Edition and K-ABC for young black and white children from low income families. Journal of Psychoeducational Assessment, 8, 139-149.

Laurent, J., Swerdlik, M., & Ryburn, M. (1992). Review of validity research on the Stanford-Binet Intelligence Scale: Fourth Edition. Psychological Assessment, 4(1), 102-112.

Lukens, J. (1988). Comparison of the Fourth Edition and the L-M Edition of the Stanford-Binet used with mentally retarded persons. Journal of School Psychology, 26, 87-89.

Lukens, J. (1990). Stanford-Binet, Fourth Edition and the WISC-R for children in the lower range of intelligence. Perceptual and Motor Skills, 70, 819-822.

McCabe, D., & Hilmo, J. (1985). Pictures speak louder than test scores. Academic Therapy, 20(3), 333-338.

McCall, V.W., Yates, B., Hendricks, S., Turner, K., & McNabb, B. (1989). Comparison between the Stanford-Binet: L-M and the Stanford-Binet: Fourth Edition with a group of gifted children. Contemporary Educational Psychology, 14, 93-96.

McGilligan, P.R., Yater, A.C., & Huesing, R. (1971). Goodenough-Harris drawing test reliabilities. Psychology in the Schools, 8, 359-362.

Naglieri, J.A. (1988). Manual for Draw a Person: A quantitative scoring system. San Antonio, TX: Psychological Corporation.

Olivier, K., & Barclay, A. (1967). Stanford-Binet and Goodenough-Harris test performances of head start children. Psychological Reports, 20, 1175-1179.

Ownby, R.L., & Carmin, C.N. (1988). Confirmatory factor analyses of the Stanford-Binet Intelligence Scale, Fourth Edition. Journal of Psychoeducational Assessment, 6, 331-340.

Phelps, L. (1989). Comparison of scores for intellectually gifted students on the WISC-R and the Fourth Edition of the Stanford-Binet. Psychology in the Schools, 26, 125-129.

Phelps, L., Bell, M.C., & Scott, M.J. (1988). Correlations between the Stanford-Binet: Fourth Edition and the WISC-R with a learning disabled population. Psychology in the Schools, 25, 380-382.

Pihl, R.O., & Nimrod, G. (1976). The reliability and validity of the Draw-a-Person test in IQ and personality assessment. Journal of Clinical Psychology, 32, 470-472.

Prewett, P.N., Bardos, A.N., & Naglieri, J.A. (1989). Assessment of mentally retarded children with the Matrix Analogies Test-Short Form, Draw a Person: A quantitative scoring system, and the Kaufman Test of Educational Achievement. Psychology in the Schools, 26, 254-260.

Ritter, D.R., Duffey, J.B., & Fischman, R. (1974). Comparisons of the intellectual estimates of the Draw-a-Person test, Peabody Picture Vocabulary test, and Stanford-Binet (L-M) for kindergarten children. Psychology in the Schools, 11, 412-415.

Rothlisberg, B.A. (1987). Comparing the Stanford-Binet, Fourth Edition to the WISC-R: A concurrent validity study. Journal of School Psychology, 25, 193-196.

Rothlisberg, B.A. (1990). The relation of the Stanford-Binet: Fourth Edition to measures of achievement: A concurrent validity study. Psychology in the Schools, 27, 120-125.

Sattler, J.M. (1988). Assessment of children. (3rd ed.) San Diego, CA: Author.

Scott, L.H. (1981). Measuring intelligence with the Goodenough-Harris Drawing test. Psychological Bulletin, 89(3), 483-505.

Strommen, E.F., & Smith, J.K. (1987). Internal consistency and bias considerations of the Goodenough-Harris Draw-a-Person test. Educational and Psychological Measurement, 47, 731-736.

Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). The Stanford-Binet Intelligence Scale, Fourth Edition-technical manual. Chicago, IL: Riverside.

Thorndike, R.M. (1990). Origins of intelligence and its measurement. Journal of Psychoeducational Assessment, 8, 223- 230.

Thorndike, R.M. (1990). Would the real factors of the Stanford-Binet Fourth Edition please come forward? Journal of Psychoeducational Assessment, 8, 412-435.

Tomes, R. & Heilbuth, L. (1991). Children's picture drawing, cognitive functioning, and neuromotor development. Unpublished data. Research funded by grants from the OSU Office of the Assistant Vice President for Research, the College of Education, and the College of Home Economics.

Tramill, J.L., Edwards, R.P., & Tramill, J.K. (1980). Comparison of the Goodenough-Harris Drawing test and the WISC-R for children experiencing academic difficulties. Perceptual and Motor Skills, 50, 543-546.

Weinberg, R.A. (1989). Intelligence and IQ: Landmark issues and great debates. American Psychologist, 44(2), 98-104.

White, T.H. (1979). Correlations among the WISC-R, PIAT, and DAM. Psychology in the Schools, 16, 497-501.

Wisniewski, J.J., & Naglieri, J.A. (1989). Validity of the Draw a Person: A quantitative scoring system with the WISC-R. Journal of Pychoeducational Assessment, 7, 346-351.

VITA

Sheila Marie Kraemer

Candidate for the Degree of

Master of Science

Thesis: COMPARISON OF THE STANFORD-BINET INTELLIGENCE SCALE WITH THE GOODENOUGH-HARRIS DRAW A PERSON AND THE DRAW A PERSON: A QUANTITATIVE SCORING SYSTEM

Major Field: Family Relations and Child Development

Biographical:

Personal Data: Born in Ada, Oklahoma, November 1, 1968, the daughter of Larry and Emma Heard.

Education: Graduated from Roff High School, Oklahoma, in May 1986: received Bachelor of Science Degree in Family Relations and Child Development from Oklahoma State University, Stillwater, Oklahoma, in May 1991; completed requirements for Master of Science degree at Oklahoma State University in May, 1993.

Professional Experience: Research Assistant, Oklahoma State University, 1991-1993; Teacher-Research Assistant, Early Intervention Program, Oklahoma State University Child Development Laboratory, 1991; Teaching Assistant, Oklahoma State University Child Development Laboratory, five year-old program, 1991.

Professional Affiliations: Society for Research in Child Development, American Psychological Association, Omicron Nu Honor Society, Phi Kappa Phi Honor Society.